

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-319497

(43)Date of publication of application : 08.12.1995

(51)Int.Cl.

G10L 5/04

(21)Application number : 06-108761

(71)Applicant : N T T DATA TSUSHIN KK

(22)Date of filing : 23.05.1994

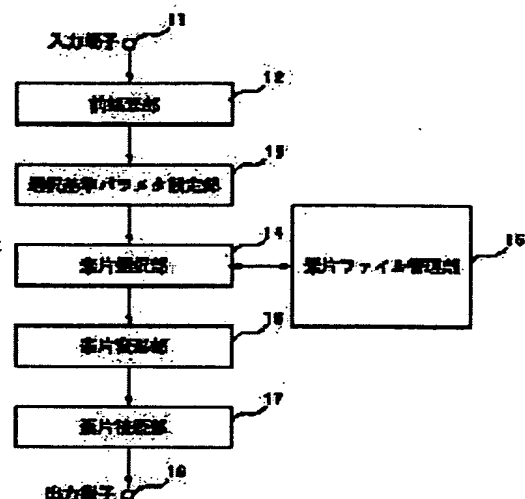
(72)Inventor : HAYASHI KEIJI  
MURAKAMI NORIYA

## (54) VOICE SYNTHESIS DEVICE

### (57)Abstract

**PURPOSE:** To provide a voice synthesis device attaining the reducing of the time for processing by reducing the storage quantity of natural voice data required for a voice synthesis while maintaining the nativeness of a voice.

**CONSTITUTION:** An input character string is divided into phonemic units in a pre-processing part 12, rhythm parameters (selection reference parameters) becoming selection references of optimum element pieces are set in a selection reference parameter setting part 13. In an element piece selecting part 14, whether waveform element pieces corresponding to respective phonemic units are present or not is retrieved and when a waveform piece with which an error with selection parameters becomes minimum is selected. On the other hand, when the waveform element piece is not present, the single syllable corresponding to the phonemic unit is selected and extracted from a single syllable file by regarding the single syllable as an optimum element piece. Then, after rhythm parameters of selected waveform element pieces or selected single syllables are deformed with an element piece deforming part 16, the synthesized voice corresponding to the input character string is obtained by combining them successively.



(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開平7-319497

(43)公開日 平成7年(1995)12月8日

(51)Int.Cl.<sup>6</sup>

G 1 0 L 5/04

識別記号

F

庁内整理番号

F I

技術表示箇所

審査請求 未請求 請求項の数 3 O L (全 6 頁)

(21)出願番号 特願平6-108761

(22)出願日 平成6年(1994)5月23日

(71)出願人 000102728

エヌ・ティ・ティ・データ通信株式会社  
東京都江東区豊洲三丁目3番3号

(72)発明者 林 慶士

東京都江東区豊洲三丁目3番3号 エヌ・  
ティ・ティ・データ通信株式会社内

(72)発明者 村上 憲也

東京都江東区豊洲三丁目3番3号 エヌ・  
ティ・ティ・データ通信株式会社内

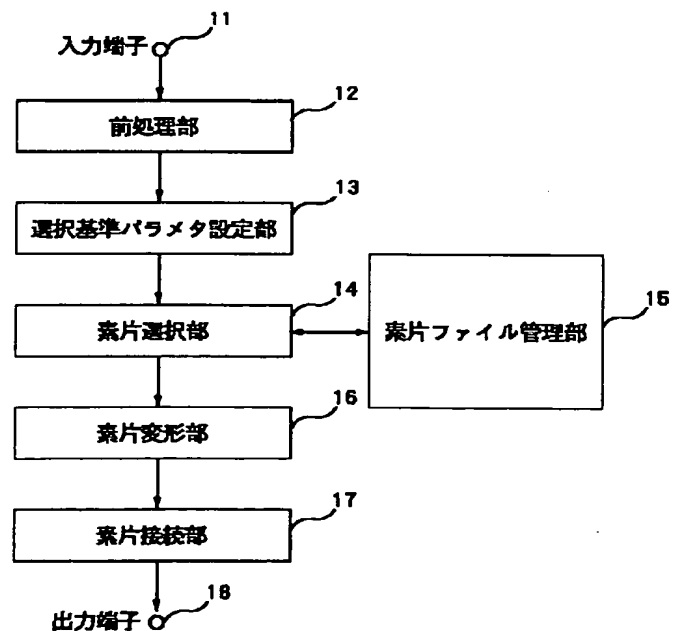
(74)代理人 弁理士 鈴木 正剛

(54)【発明の名称】 音声合成装置

(57)【要約】

【目的】 音声の自然性を確保しつつ音声合成に必要な自然音声データの格納量を減らして処理時間の短縮を図る構成の音声合成装置を提供する。

【構成】 入力文字列を前処理部12で音韻単位に分割し、選択基準パラメタ設定部13で最適素片の選択基準となる韻律パラメタ(選択基準パラメタ)を設定する。素片選択部14では、素片ファイル管理部15を介して各音韻単位に対応する波形素片が存在するか否かを検索し、存在するときは選択基準パラメタとの誤差が最小となる波形素片を選択抽出する。他方、波形素片が存在しないときは、当該音韻単位に対応する単音節を最適素片とみなして単音節ファイルから選択抽出する。選択した波形素片又は単音節の韻律パラメタを素片変形部16で変形した後、素片接続部で順次結合して入力文字列に対応する合成音声を得る。



## 【特許請求の範囲】

【請求項1】 音声に対応する入力文字列を音韻単位に分割する前処理部と、  
自然音声より切り出した複数の波形素片と各波形素片の韻律パラメタとを格納する素片情報格納手段と、  
複数の単音節と各単音節の韻律パラメタとを格納する単音節情報格納手段と、  
前記分割された音韻単位に対応する波形素片が前記素片情報格納手段に存在するか否かを検索し、存在するときは予め素片選択基準として定めた韻律パラメタとの誤差が最小となる韻律パラメタに対応する波形素片を選択抽出し、他方、存在しないときは当該音韻単位に対応する単音節を前記単音節情報格納手段より選択抽出する素片選択手段と、  
抽出された波形素片又は単音節を前記入力文字列の順に接続して合成音声を生成する素片接続部と、  
を少なくとも有することを特徴とする音声合成装置。

【請求項2】 請求項1記載の音声合成装置において、前記素片選択手段で選択抽出した波形素片又は単音節の韻律パラメタを変形せしめて前記素片選択基準として定めた韻律パラメタとの誤差を零値に近づける素片変形部を設け、この変形された韻律パラメタの波形素片又は単音節を前記素片接続部に導く構成としたことを特徴とする音声合成装置。

【請求項3】 請求項1又は2記載の音声合成装置において、前記誤差は、前記選択基準として定めた韻律パラメタと抽出された個々の韻律パラメタとの差分を各韻律パラメタの変動幅で除した値の2乗の和であることを特徴とする音声合成装置。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】本発明は音声合成装置に関し、特に、合成パラメタとして自然音声から切り出して得た波形素片と韻律パラメタが既知の単音節とを用いて、入力文字列から合成音声を生成する音声合成装置に関する。

## 【0002】

【従来の技術】図6は、従来の一般的な音声合成装置の機能ブロック図の一例を示す図であり、61は入力端子、62は前処理部、63は選択基準パラメタ設定部、64は素片選択部、651は素片パラメタテーブル、652は素片ファイル、66は素片変形部、67は素片接続部、68は出力端子である。

【0003】この構成の音声合成装置において、音素記号及びアクセント記号からなる入力文字列は、入力端子61から入力された後に前処理部62において音韻単位に分割される。選択基準パラメタ設定部63では、分割された音韻単位とアクセント記号とから、選択基準パラメタ（平均ピッチ周波数・ピッチ傾斜・時間長・平均パワー）を設定し、素片選択部64に出力する。なお、上記

選択基準パラメタは、合成パラメタである波形素片の選択基準として用いる韻律パラメタである。

【0004】素片ファイル652には、小説や随筆等の自然音声から切り出して得た複数の波形素片が格納されており、素片パラメタテーブル651には、各波形素片の韻律パラメタが格納されている。素片選択部64は、上記選択基準パラメタに基づいて素片パラメタテーブル651を検索し、音韻接続に対する最適素片をそれぞれ素片ファイル652から選択するものである。素片変形部66では、素片選択部64で選択された最適素片を前記選択基準パラメタに一致するよう適宜素片の変形処理を施す。素片接続部67では、変形された素片をそれぞれ入力文字列にしたがって接続し、出力端子68を通じて出力する。

## 【0005】

【発明が解決しようとする課題】ところで、上述の音声合成処理の際に、音韻単位に対応する波形素片が素片ファイル652に存在しないと、当該音韻単位についての音声が生産されない。従って、自然な合成音声を得るためには、合成対象となる文字列の語彙に対応する波形素片を少なくとも1つは用意しておく必要がある。しかしながら、任意度の高い語彙について合成音声処理を行う場合に、対応する全ての波形素片を予め素片ファイル652に用意し、その韻律パラメタを解析しておくことは困難であり、音声品質の向上に一定の限界があった。素片ファイル652に新たに波形素片を追加することもできるが、いずれの場合もかなり大量の波形素片を収集する必要があり、多大な労力を要する問題があった。

【0006】また、上述のように素片ファイル652に大量の波形素片を格納すると必然的にファイル容量が大きくなる。そのため、合成処理に要する時間がそのファイル容量に比例して増大するという問題もあった。

【0007】本発明は上記問題点を解消し、音声合成に必要な自然音声データの量を減らして処理時間の短縮を図ることができ、しかも自然な合成音声を得られる構成の音声合成装置を提供することを目的とする。

## 【0008】

【課題を解決するための手段】本発明が提供する音声合成装置は、少なくとも、音声に対応する入力文字列を音韻単位に分割する前処理部と、自然音声より切り出した複数の波形素片と各波形素片の韻律パラメタとを格納してなる素片情報格納手段と、複数の単音節と各単音節の韻律パラメタとを格納してなる単音節情報格納手段と、前記分割された音韻単位に対応する波形素片が前記素片情報格納手段に存在するか否かを検索し、存在するときは予め素片選択基準として定めた韻律パラメタとの誤差が最小となる韻律パラメタに対応する波形素片を選択抽出し、他方、存在しないときは当該音韻単位に対応する単音節を前記単音節情報格納手段より選択抽出する素片選択手段と、抽出された波形素片又は単音節を前記入力

文字列の順に接続して合成音声を生成する素片接続部と、を有することを特徴とする。つまり、波形素片及び単音節の2種のパラメタを複合的な合成パラメタとして用いる。

【0009】本発明が提供する他の構成の音声合成装置は、上記音声合成装置において、更に、前記素片選択手段で選択抽出した波形素片又は単音節の韻律パラメタを変形せしめて前記素片選択基準として定めた韻律パラメタとの誤差を零値に近づける素片変形部を設け、この変形された韻律パラメタの波形素片又は単音節を前記素片接続部に導く構成としたことを特徴とする。

【0010】なお、前記誤差は、好ましくは、前記選択基準として定めた韻律パラメタと抽出された個々の韻律パラメタとの差分を各韻律パラメタの変動幅で除した値の2乗の和とする。

【0011】

【作用】本発明の音声合成装置では、入力文字列を前処理部で音韻単位に分割し、個々の音韻単位に対する波形素片の選択基準となる韻律パラメタ（選択基準パラメタ）を設定する。この選択基準パラメタは、例えば所望の平均ピッチ周波数、ピッチ傾斜、時間長、平均パワーである。素片選択手段においては、まず、各音韻単位に対応する波形素片が素片情報格納手段に存在するか否かを検索する。存在するときは、これを該素片情報格納手段より選択抽出する。複数存在するときは、例えば選択基準パラメタと各波形素片の韻律パラメタとの間で2乗誤差を計算し、その2乗誤差が最小となる波形素片を選択抽出する。他方、波形素片が存在しないときは、当該音韻単位に対応する単音節を最適素片とみなして単音節情報格納手段より選択抽出する。このようにして選択した波形素片又は単音節を、素片接続手段が入力文字列に従って音韻単位毎に順次結合し、合成音声を生成する。

【0012】本発明の他の構成の音声合成装置では、上記素片選択手段で選択抽出した波形素片又は単音節の韻律パラメタを変形せしめ、素片選択基準として定めた韻律パラメタとの誤差を零値に近づける。そして変形された韻律パラメタの波形素片又は単音節を素片接続部に導く。これにより合成音声の韻律が調整される。

【0013】

【実施例】以下、図面を参照して本発明の実施例を詳細に説明する。図1は、本発明の一実施例に係る音声合成装置の機能ブロック図であり、11は入力端子、12は前処理部、13は選択基準パラメタ設定部、14は素片選択部、15は素片ファイル管理部、16は素片変形部、17は素片接続部、18は出力端子である。素片選択部14と素片ファイル管理部とで本発明の素片選択手段を構成する。素片ファイル管理部15を除く各部の機能は、基本的には図6に示した従来装置のものと同様なので、その詳細な説明は省略する。

【0014】図2に、本実施例の素片ファイル管理部1

5の構成例を示す。この素片ファイル管理部15は、音韻情報管理テーブル22、素片ファイル241、素片パラメタテーブル231、単音節パラメタテーブル232、単音節ファイル242、及び、図示を省略したファイル制御部（CPU）から成る。なお、CPU、素片ファイル241、及び、素片パラメタテーブル231を含んで本発明の素片情報格納手段を構成し、同様に、CPU、単音節ファイル242、及び、単音節パラメタテーブル232を含んで本発明の単音節情報格納手段を構成する。

【0015】素片ファイル241には、音韻単位に、予め音韻環境等を考慮して単語・文章から切り出された素片が複数個格納されている。これに対し、単音節ファイル242には、50音や連母音等の単音節がそのまま格納されている。これら両ファイル241、242に格納された波形素片及び単音節が、それぞれ後述の複合的な合成パラメタとして用いられる。音韻情報管理テーブル22には、個々の音韻単位に属する素片ファイル241中の波形素片数が、例えば数値データとして記述されている。また、素片パラメタテーブル231には、各波形素片の韻律パラメタがそれぞれ格納され、単音節パラメタテーブル232には、各単音節の韻律パラメタが格納されている。これらテーブル231、232の内容例を図3、図4に示す。

【0016】図3は、／k a／に関する素片パラメタテーブル231の内容例であり、30はテーブル名、31はファイル番号、32は素片抽出環境、33は平均ピッチ周波数、34はピッチ傾斜、35は継続時間、36は平均パワーである。テーブル名30は、例えば素片抽出環境32のインデックスとして機能させる。「#k a-」は、自然音声（単語・文章）の先頭にある音韻が／か／で、アクセントの付与されていないものを意味する。ファイル番号31は、素片ファイル241中のファイル識別符号であり、素片選択部14において選択の対象となる。素片抽出環境32は、波形素片の抽出可能な単語を示すものである。図示の例では、「#k a-」の波形素片が「#k a k a e r u #」等の単語から抽出可能であることを示している。各素片抽出環境に対応する韻律パラメタ33～36は、波形切り出しの際に解析された値であり、固定的な数値データである。

【0017】また、図4は、単音節パラメタテーブル232の内容例であり、41はファイル番号、42は単音節、43は平均ピッチ周波数、44はピッチ傾斜、45は継続時間、46は平均パワーである。ファイル番号41は、単音節ファイル242中のファイル識別符号であり、上記ファイル番号31と同様、素片選択部14において選択の対象となる。単音節42は、単音節ファイル242中に格納された単音節に対応するものである。図示の例では、a, i, ... wa等、110個の単音節が単音節ファイル242に格納されていることを示して

いる。この単音節の種類はできるだけ多い方が好ましい。なお、各単音節に対応する韻律パラメタ43～46は、予め解析された固定的な数値データである。

【0018】次に、上記構成の音声合成装置における音声合成処理の内容を説明する。音韻記号及びアクセント記号から成る入力文字列は、入力端子11から入力された後に、前処理部12において音韻単位に分割される。本実施例における音韻単位とは、／a／や／k a／などの音韻の他に、素片接続の滑らかさを考慮して／a i／などの連母音や、／a N／などの複合音節を含むものとする。選択基準パラメタ設定部13では、分割された音韻単位に対応する最適な波形素片（最適素片）の選択基準となる韻律パラメタ、即ち選択基準パラメタを設定する。この実施例では、選択基準パラメタとして上述の平均ピッチ周波数、ピッチ傾斜、時間長、平均パワの全てを使用するが、この選択基準パラメタは、合成音声の韻律条件に応じて適宜変更しても良い。

【0019】素片選択部14では、素片ファイル241に属する波形素片または単音節ファイル242に属する単音節の各韻律パラメタと上記選択基準パラメタとを用いて音韻単位毎の最適素片をそれぞれ選択する。この素片選択部14の詳細な動作を、図5を参照して具体的に説明する。ここで図5は、本発明における素片選択部14の動作原理を示すフローチャートであり、Sは、各処理ステップを表す。

【0020】まず、音韻情報管理テーブル22を参照して音韻単位毎の波形素片数を求める（S31）。そして、この求めた波形素片数によって以下のように選択処理の内容を決定する（S32）。波形素片数が0、つまり存在しないときは、当該音韻単位に対応する単音節を最適素片とみなして選択する（S331）。例えば、音韻単位／d i／に対応する波形素片が存在しないときは、単音節パラメタテーブル232を検索し、ファイル番号110（図4参照）の単音節d iを自動的に最適素片とみなす。他方、波形素片数が0でないときは、上記選択基準パラメタと当該波形素片の韻律パラメタとを用いて最適素片を選択する（S332）。波形素片数が1の場合は、その波形素片を最適素片とする。波形素片が2以上の場合の選択手法は、任意のものをを用いて良いが、例えば、選択基準パラメタと各韻律パラメタとの間で2乗誤差を計算し、その2乗誤差が最小となるものを最適素片とすれば、より選択基準パラメタに近い韻律パラメタの最適素片が得られる利点がある。このようにして選択した最適素片に対応するファイル及びその韻律パラメタを、素片変形部に出力する（S34）。

【0021】例えば、図3におけるファイル番号0741に属する波形素片が最適素片とされた場合には、当該ファイル番号に対応するファイルを素片ファイル241より抽出して出力するとともに、その韻律パラメタ33～36を出力する。同様に、図4におけるファイル番号

110に属する単音節d iが最適素片とみなされた場合には、当該ファイル番号に対応するファイルを単音節ファイル242より抽出して出力するとともに、その韻律パラメタ43～46を出力する。

【0022】素片変形部16では、素片選択部14から導かれたファイルの韻律パラメタを、上記選択基準パラメタに近づくように変形処理を施す。この変形処理は、ファイル内容が波形素片の場合は従来の一般的な手法を用いることができる。他方、単音節の場合には、例えば、その韻律パラメタの時間長に対応する選択基準パラメタの時間長に近づくように間引き又は補間する処理となる。他の韻律パラメタについてもピッチ制御や周波数制御等によって変形処理が可能である。このように単音節の韻律パラメタについても変形処理を行うのは、単純に単音節ファイル242から該当する単音節を抽出して接続しただけでは合成音声の自然性向上が図れないためであり、また、波形素片との代替性をより完全に担保するためでもある。

【0023】素片接続部17では、前記素片変形部16で変形された波形素片又は単音節を、入力文字列に従って音韻単位毎に順次結合することによって合成音声を作成する。以上のような一連の処理を経て、出力端子18には、入力文字列に対応した、抜けの無い自然な合成音声出力される。

【0024】このように、本実施例によれば、分割された音韻単位に対応する波形素片が存在しない場合でも、対応する適切な単音節を選択して音声合成を行うことができる。従って、従来のように、全ての音韻単位に対して波形素片を予め用意する必要がなく、少ない自然音声データからでも音声合成を行うことができる。

#### 【0025】

【発明の効果】以上の説明から明らかなように、本発明の音声合成装置は、波形素片及び単音節の2種のパラメタを複合的な合成パラメタとして用い、音韻単位に対応する波形素片が素片情報格納手段に存在するときは、予め素片選択基準として定めた韻律パラメタとの誤差が最小となる韻律パラメタに対応する波形素片を選択抽出し、他方、存在しないときは当該音韻単位に対応する単音節を単音節情報格納手段より選択抽出し、抽出された波形素片又は単音節を入力文字列の順に接続する構成なので、合成対象となる語彙が増加した場合でも、語彙に対応する波形素片を新たに追加することなく、予め作成した自然音声データ、即ち波形素片と単音節、及びその韻律パラメタのみを用いて、抜けのない自然な合成音声を作成することができる。しかも、上述のように自然音声データの追加を必要としないので、ファイル容量を小さくすることができる。これにより、簡素で低コスト、且つ、音声合成処理時間が短縮されて一定時間内に高品質な合成音声を作成する構成の音声合成装置を実現することができる。

【0026】また、本発明の他の構成に係る音声合成装置は、上記素片選択部で選択された波形素片又は単音節の韻律パラメタを変形して素片選択の基準となる韻律パラメタに近付ける素片変形部を設け、この変形された波形素片等を接続して合成音声を生成するようにしたので、所望の韻律の合成音声を高速に生成することが可能となる。

【図面の簡単な説明】

【図1】本発明の一実施例に係る音声合成装置の機能ブロック図。

【図2】本実施例の音声合成装置における素片ファイル管理部の構成図。

【図3】上記素片ファイル管理部を構成する素片パラメタテーブルの内容説明図。

【図4】上記素片ファイル管理部を構成する単音節パラメタテーブルの内容説明図。

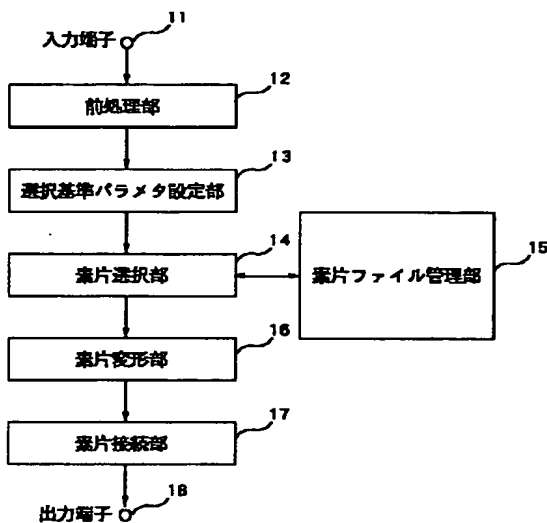
【図5】本実施例の音声合成装置における素片選択部の動作を示すフローチャート。

【図6】従来の音声合成装置の機能ブロック図

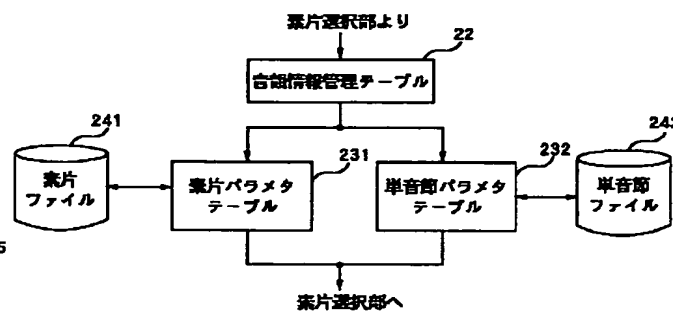
【符号の説明】

- 1 2 前処理部
- 1 3 選択基準パラメタ設定部
- 1 4 素片選択部
- 1 5 素片ファイル管理部
- 1 6 素片変形部
- 1 7 素片接続部
- 2 2 音韻情報管理テーブル
- 2 3 1 素片パラメタテーブル
- 2 3 2 単音節パラメタテーブル
- 2 4 1 素片ファイル
- 2 4 2 単音節ファイル

【図1】



【図2】



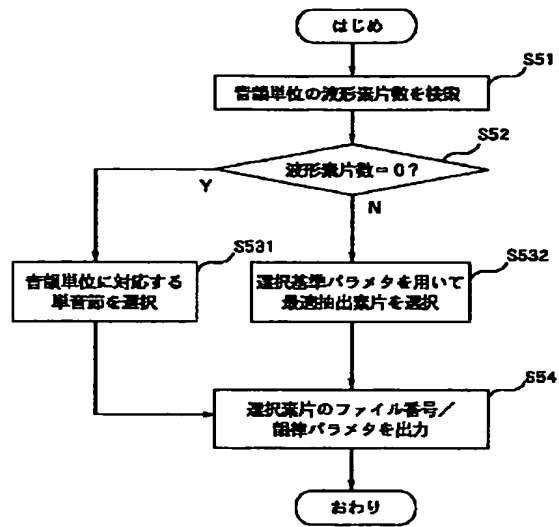
【図4】

ファイル番号	単音節	平均ピッチ 周波数 (Hz)	ピッチ 傾斜	継続時間 (ms)	平均パワー
001	a	150.00	0.20	170.00	3100
002	i	125.87	0.70	90.00	1054
⋮					
050	wa	185.00	1.60	195.00	8800
⋮					
110	di	140.25	0.87	85.75	884

【図3】

ファイル番号	素片抽出環境	平均ピッチ 周波数 (Hz)	ピッチ 傾斜	継続時間 (ms)	平均パワー
0741	# kakaeru #	131.00	0.00	70.08	1164
0742	# kakageru #	129.87	0.70	55.08	1054
⋮					
0745	# kakaru #	138.00	1.60	45.08	946
⋮					
0748	# kogami #	131.25	0.87	63.08	884

【図 5】



【図 6】

